



## **A simplified DNA extraction protocol for unsorted bulk arthropod samples that maintains exoskeletal integrity**

Nielsen, Martin; Gilbert, M. Thomas P.; Pape, Thomas; Bohmann, Kristine

*Published in:*  
Environmental DNA

*DOI:*  
[10.1002/edn3.16](https://doi.org/10.1002/edn3.16)

*Publication date:*  
2019

*Document version*  
Publisher's PDF, also known as Version of record

*Document license:*  
[CC BY](#)

*Citation for published version (APA):*  
Nielsen, M., Gilbert, M. T. P., Pape, T., & Bohmann, K. (2019). A simplified DNA extraction protocol for unsorted bulk arthropod samples that maintains exoskeletal integrity. *Environmental DNA*, 1(2), 144-154.  
<https://doi.org/10.1002/edn3.16>

## ORIGINAL ARTICLE

# A simplified DNA extraction protocol for unsorted bulk arthropod samples that maintains exoskeletal integrity

Martin Nielsen<sup>1,2</sup>  | M. Thomas P. Gilbert<sup>1,2,3</sup>  | Thomas Pape<sup>1</sup>  |  
Kristine Bohmann<sup>1,2,4</sup> 

<sup>1</sup>Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

<sup>2</sup>Section for Evolutionary Genomics, University of Copenhagen, Copenhagen, Denmark

<sup>3</sup>NTNU University Museum, Trondheim, Norway

<sup>4</sup>School of Biological Sciences, Norwich Research Park, University of East Anglia, Norwich, UK

## Correspondence

Martin Nielsen, Section for Evolutionary Genomics, University of Copenhagen, Øster Farimagsgade 5a, opg. 7, 2. sal, 1307 Copenhagen, Denmark.  
Email: mnielsen@science.ku.dk

## Funding information

EvoGenomics MSc grant; Independent Research Fund Denmark, Grant/Award Number: DFF 5051-00140; Dr. Bøje Benzons Support grant; Carlsberg Foundation, Grant/Award Number: 2012\_01\_0504; Aage V. Jensens Fonde and ERC Consolidator Grant

## Abstract

High-throughput DNA sequencing offers an efficient tool for assessing the taxonomic content of bulk arthropod samples. Many current DNA extraction protocols however require extensive handling of samples, like specimen-based DNA extractions, or sorting of samples and are thus unsuitable for large scale studies. Furthermore, protocols often include homogenization and thus imply partial or complete destruction of the sample constituents. The aim of this study was therefore to investigate steps related both to sample pre-processing and DNA extraction of unsorted bulk arthropod samples and explore possibilities for simplifying sample processing and thus increase sample handling efficiency without losing taxonomic information. Using mock bulk arthropod samples, we compare laboratory steps related to DNA extraction and semi-automatic handling. Specifically, we (a) assess whether aliquots of digest buffer from bulk arthropod samples adequately describe the community composition; (b) compare a non-destructive and a destructive DNA extraction method; (c) compare a phenol/chloroform inhibitor removal method with the exclusion of the same; and (d) compare manual purification to automated DNA purification on a QIAcube laboratory robot. Using DNA metabarcoding and mock bulk arthropod samples, we show that it is possible to efficiently process unsorted arthropod bulk samples with a non-destructive DNA extraction approach. We found that homogenizing samples yielded more DNA but also generally produced more inconsistent results when compared to non-destructive extraction. When assessing the recovered taxonomic content of samples (operational taxonomic units, OTUs), intact samples performed at least comparable to, if not better, than homogenized samples. Additionally, we show that sample processing can be further simplified from using a defined volume of digest, no phenol/chloroform purification and automated DNA purification. This approach can be a way to process hundreds of unsorted bulk samples effectively, consistently and with a minimum loss of valuable morphologic information.

## KEYWORDS

arthropod diversity, community samples, DNA metabarcoding, high-throughput sequencing, insect diversity

## 1 | INTRODUCTION

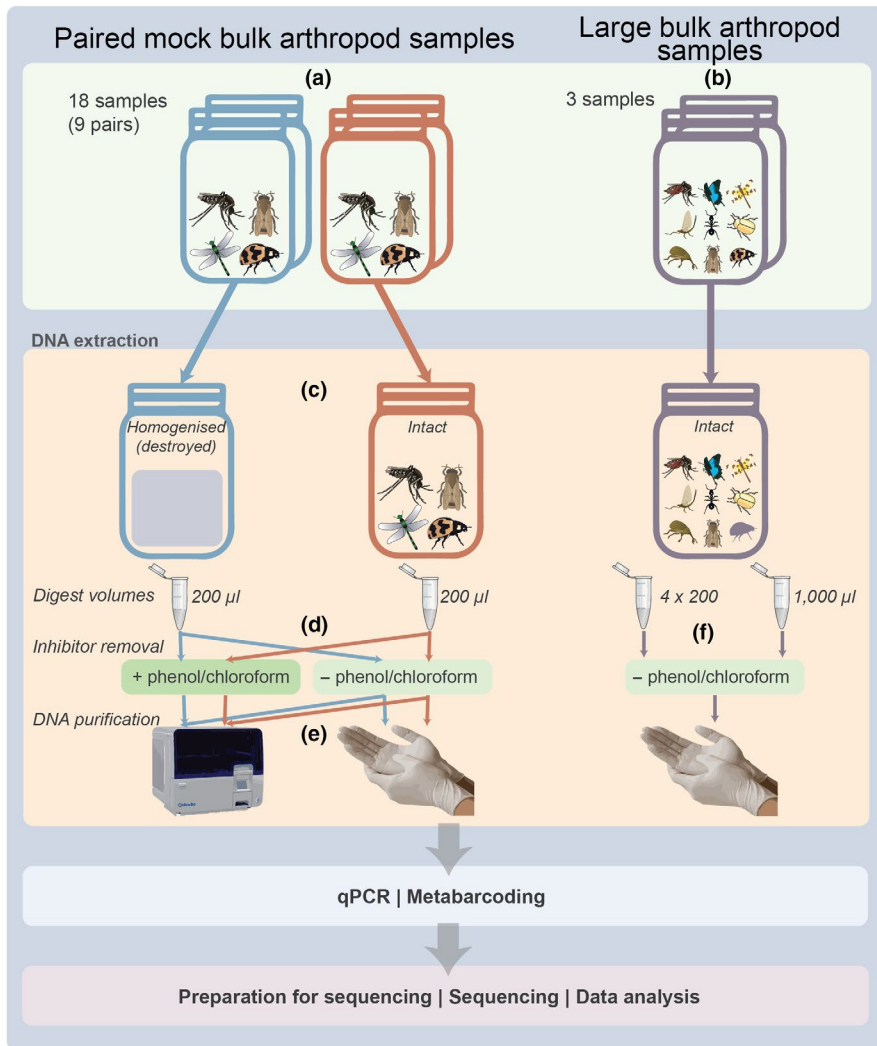
Arthropods greatly outnumber all other eukaryotic taxa in terms of species (Mora, Tittensor, Adl, Simpson, & Worm, 2011), and they have a profound influence on human well-being, ranging from vectors of deadly diseases to providing essential ecosystem services, such as pollination and pest control (e.g. reviewed in Noriega et al., 2018). However, the immense number of arthropod species has traditionally represented an enormous challenge for ecological studies, monitoring programs, and biodiversity assessment analyses that rely on bulk arthropod samples. Such studies can include hundreds of bulk samples each containing numerous species of which many might even be undescribed. Morphology-based identification of specimens in such bulk arthropod samples is a daunting task that can require thousands of working hours and a wide range of taxonomic expertise (Basset et al., 2012). One solution has been to restrict the focus, and thereby the taxonomic identifications, to indicator groups such as butterflies (e.g. Thomas et al., 2004) or beetles (e.g. Rainio & Niemelä, 2003), which while faster, results in the loss of much of the information contained within the samples and adds a taxonomic bias. DNA-based methods have been suggested as an alternative means with which to improve the speed and economic cost of biodiversity assessments in general (Baird & Hajibabaei, 2012), and in the case of arthropods, such approaches offer the potential to move from looking at just a few indicator species to assessing entire communities. Among the DNA-based methods, metabarcoding is currently the methodology of choice, making it possible to identify multiple taxa with high-throughput sequencing (HTS) of DNA mini-barcodes obtained from PCR amplification (Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012). Studies have shown that metabarcoding can be used to identify arthropod taxa in bulk arthropod samples (e.g. Kocher et al., 2017; Liu et al., 2013; Morinière et al., 2016), and for example be applied to assess how arthropod diversity changes to different land-uses (Beng et al., 2016). Metabarcoding has also been used to assess how invertebrates might leak DNA to their preservative ethanol which then potentially can serve as a mean to obtain sample diversity information (Erdozain et al., 2019; Hajibabaei, Spall, Shokralla, & Konynenburg, 2012). As an alternative to metabarcoding, shotgun sequencing, where the total DNA is sequenced, has been explored, for example to identify the species composition of insect communities (Zhou et al., 2013) and to elucidate beetle phylogenies (Crampton-Platt et al., 2015; Gillett et al., 2014).

A key element of any DNA-based study of bulk arthropod samples is the DNA extraction, which often includes physical homogenization of individual bulk samples into "insect soups" before DNA extraction, losing valuable morphological information (Gibson et al., 2014; Gillett et al., 2014; Morinière et al., 2016; Yu et al., 2012). Samples often undergo additional processing during the DNA extraction process, for example sorting into individual specimens to amplify their individual DNA barcodes

(Wang, Srivathsan, Foo, Yamane, & Meier, 2018). To account for biomass differences, specimens within bulk samples have been sorted into size-groups (Elbrecht, Peinert, & Leese, 2017) or taxonomic groups (Morinière et al., 2016). Bulk samples have also been sorted out for individual DNA extractions, which were then pooled together to mimic community samples (Brandon-Mong et al., 2015; Crampton-Platt et al., 2015; Yu et al., 2012). While the rationale behind each of these approaches is to maximize the recovered taxonomic information, the labor incurred renders them unrealistic for large-scale studies with hundreds of bulk arthropod samples.

Despite the central role that DNA extraction takes, few studies have addressed how this can be optimized for handling large numbers of bulk samples where contamination risk, consistency, inhibitor levels, and handling time are important factors. For example, avoiding sorting or homogenization requires large volumes of digestion buffer to extract DNA from bulk arthropod samples with hundreds of individuals. This further opens the question as to whether one needs to purify all the digest in order to reconstruct the taxonomic composition of the sample, or if purifying a smaller volume of the digest can provide representative information. Additionally, the widespread initial physical homogenization of the bulk arthropod samples, with the associated loss of morphological information, may be circumvented. This would save time, reduce contamination risk, and preserve valuable morphological information. Previous studies have shown that it is possible to extract DNA by leaching DNA from individual insects while keeping their exoskeletal structures intact (e.g. Gilbert, Moore, Melchior, & Worobey, 2007), thus raising the obvious question as to whether such non-destructive extraction methods can be applied successfully to bulk arthropod samples. Additionally, DNA extracted from bulk arthropod samples can contain PCR inhibitors that may negatively affect the amplification success (Murray, Coghlan, & Bunce, 2015; Schrader, Schielke, Ellerbroek, & John, 2012) and which therefore require removal through labor-intensive purification, raising the question as to whether such methods add any benefit to the result of bulk arthropod studies. Lastly, there is an obvious need to automate extractions of bulk arthropod samples as this will make processing large numbers of samples more feasible, but the effectiveness and reliability of automation needs validation and quantified comparisons to manual approaches.

In this study we aim at answering the above-mentioned questions. First, we compare a non-destructive DNA extraction method to a destructive method. Second, we assess whether large quantities of digest obtained from bulk arthropod samples need to be purified, or if purifying a subset of the digest is sufficient to recover taxonomic information of bulk arthropod samples. Third, we assess the need to use phenol/chloroform to remove inhibitors from extracts obtained from bulk arthropod samples. Lastly, we evaluate whether automatic purification of digests on a QIAcube robot (Qiagen) performs as well as manual purification. A detailed overview of the study is given in Figure 1.



**FIGURE 1** Study overview. Twenty-one mock arthropod bulk samples falling into two categories were created: (a) Nine sample pairs (18 samples in total) in which each sample contained 11–100 individual arthropods and (b) three large samples containing 100, 510, and an unknown number of arthropods. The nine sample pairs were used to (c) compare physical homogenization with no homogenization prior to DNA extraction, (d) assess the effect of phenol/chloroform inhibitor removal, and (e) compare DNA purification on a QIAcube robot to manual purification. The three larger samples were used to (f) assess the effect of purifying different digest volumes. Samples that underwent a phenol/chloroform inhibitor removal step were only purified on a QIAcube robot. Comparisons and assessments were based on quantitative PCR and metabarcoding. Note that it was not possible to make the nine sample pairs perfectly identical

## 2 | MATERIAL AND METHODS

### 2.1 | Mock bulk arthropod samples

Twenty-one mock bulk arthropod samples were generated from material collected in Malaise traps between 2001 and 2014 in Denmark, Greenland, and Tanzania and kept in 70% ethanol at ambient temperature as part of the collections at the Natural History Museum of Denmark. The mock samples fell into two categories: (a) 18 bulk arthropod samples forming nine sample pairs containing 11–100 individual arthropod specimens; and (b) three bulk arthropod samples of varying size; one of 100 specimens, one of 510 specimens, and one large sample containing 10 ml of an unknown number of specimens. The 18 samples in the nine sample pairs were generated to be as pairwise similar as possible with regard to the number, body-sizes, and origin of specimens (Table S1). Thus, each sample pair contained the same number of specimens within the same three size categories (small, medium, and large) all originating from the same Malaise trap sample. For example: The two mock bulk arthropod samples in sample pair 2 contained 30 small, 0 medium, and 2 large-sized arthropods all originating from

the same Malaise trap sample collected in 2014 in Tanzania (Table S1). Unfortunately, it was not possible to make the nine sample pairs fully identical. Mock samples were photographed with a digital single-lens reflex camera and specimens were morphologically identified to either taxonomic order or family level based on the images. Specimens in sample pair nine and the three larger samples were not identified because of the high number of specimens. All mock samples were stored in 70% ethanol at room temperature until DNA extraction.

### 2.2 | DNA extraction

DNA extractions were carried out in a dedicated pre-PCR laboratory. Ethanol was carefully poured off the samples, and remaining ethanol was evaporated by placing the samples in an oven at 56°C until dry. Within each of the nine sample pairs, one sample was randomly chosen to be kept intact, while the other was homogenized using a small grinder covered with aluminium foil (Figure 1c). New foil was used and the grinder cleaned with bleach and ethanol between each sample. The three large samples were

kept intact. Digestion of all samples was carried out using a digestion buffer modified from (Gilbert et al., 2007) consisting of 10 mmol/L Tris-HCl (pH 8.0), 10 mmol/L NaCl, 5 mmol/L CaCl<sub>2</sub>, 2.5 mmol/L ethylene-diamine-tetra-acetic acid (EDTA), 1% sodium dodecyl sulfate (SDS), 1% proteinase K, 40 mmol/L dithiothreitol (DTT), and molecular biology grade H<sub>2</sub>O. Digestion buffer was added so that all specimens or specimen parts in a sample were covered. A negative extraction control was included alongside the digestion of homogenized samples, intact samples, and the three larger samples. After adding digestion buffer, samples were placed on a rotator in an oven at 56°C overnight (>14 hr) and briefly centrifuged to pellet arthropods. After collecting the DNA digest, 100% ethanol was added to the arthropod bulk samples to stop digestion. Ethanol was replenished with 70% ethanol after 1–2 hr where after samples were returned to the collection at the Natural History Museum of Denmark.

From each of the three larger samples, 4 × 200 µl and 1 × 1,000 µl digest were purified manually using the QiaQuick PCR purification kit (Qiagen, CA), see specification below (Figure 1f). From each of the 18 samples in the nine sample pairs (nine homogenized and nine intact) and negative extraction controls, 200 µl digest was purified with three different approaches: (a) phenol/chloroform inhibitor removal followed by automatic purification on QIAcube robot (Qiagen, Valencia, CA), see specification below; (b) only manual purification; (c) only automatic purification (Figure 1d,e). Phenol/chloroform inhibitor removal was carried out by adding 200 µl phenol/chloroform solution (25:24:1 phenol:chloroform:isoamyl alcohol, Sigma-Aldrich, St. Louis, MI) to 200 µl digest. This was mixed by resuspension, placed on a rotator for 5 min followed by centrifugation at 17,000× g for 5 min after which the supernatant was removed for later purification. Automatic purifications were carried out on a QIAcube robot (Protocol: “Purification of PCR products from 100–200 µl PCR samples”, version 1.0) with the following settings: a binding step with 5× volume of PB buffer followed by centrifugation at 12,000× g for 60 s. A wash step with 750 µl PE buffer followed by centrifugation at 12,000× g for 60 s and a subsequent dry-spin centrifugation at 12,000× g for 60 s. Lastly, DNA was eluted in 30 µl elution buffer (EB), incubated at 1 min at room temperature and followed by centrifugation at 12,000× g for 60 s. Manual purifications were carried out with the QiaQuick PCR purification kit (Qiagen) following the manufacturer's protocol, but with the same parameters as for the QIAcube robot. DNA extracts were stored at –20°C until further analyses.

## 2.3 | Quantitative PCR

All DNA extracts were quantitative PCR (qPCR) screened at neat and 1:10 dilutions using the arthropod cytochrome c oxidase subunit I mini-barcode primers ZBJ-ArtF1c and ZBJ-ArtR2c (Zeale, Butlin, Barker, Lees, & Jones, 2011), amplifying a 157 bp fragment (excluding primers). Undiluted negative controls were also included. This was performed to screen negative extraction controls for contamination, assess DNA yield and DNA purity (screening for PCR inhibitors), and determine the optimal conditions for the following tagged

PCR (Murray et al., 2015). Four positive controls (dung beetle DNA extracts) were included in each run as “inter-run calibrators” (Bustin et al., 2010). Quantitative PCRs were carried out on an Agilent Technologies Stratagene Mx3005p quantitative PCR thermocycler (Agilent Technologies, Santa Clara, CA) in 25 µl reactions consisting of 1 µl DNA template, 1 U AmpliTaq Gold, 1× Gold PCR Buffer and 2.5 mmol/L MgCl<sub>2</sub> (all from Applied Biosystems), 0.2 mmol/L deoxyribonucleotide triphosphate Mix (Invitrogen), 5× Purified bovine serum albumin (Bio Labs), 0.6 µmol/L of each primer (ZBJ-ArtF1c and ZBJ-ArtR2c (Zeale et al., 2011)), and 1 µl of SYBR Green/ROX solution (one part SYBR Green I nucleic acid gel stain (S7563; Invitrogen), four parts ROX reference dye (12223-012; Invitrogen) and 2000 parts high-grade dimethyl sulfoxide). Forward and reverse primers were 5' nucleotide tagged with 7–8 nucleotides in length (Binladen et al., 2007), see details below. Quantitative PCR amplifications were run at 95°C for 5 min, followed by 40 cycles at 95°C for 15 s, 52°C for 30 s and 72°C for 30 s, followed by a dissociation segment of 95°C for 1 min, 52°C for 30 s, and 95°C for 30 s. All negative extraction controls appeared negative.

## 2.4 | Metabarcoding

Metabarcoding was carried out on all 69 DNA sample extracts and negative extraction controls. To minimize contamination risk, all PCRs were set up in dedicated amplicon-free laboratories. All PCR amplifications were carried out on a 2720 Thermal Cycler (Applied Biosystems) using the above-mentioned arthropod primers. PCR amplifications were carried out as described for the qPCR above, but as determined by the qPCR screening, they were carried out with 1:10 dilutions of sample DNA extracts and with 35 PCR cycles. Furthermore, SYBR Green/ROX and dissociation curve were omitted while a final extension of 72°C for 7 min was included. A negative PCR control was included for every seven reactions. A set of 60 uniquely 5'-nucleotide tagged forward and 60 uniquely tagged reverse primers were used (for tag sequences, see Schnell, Bohmann, & Gilbert, 2015). Tags were 7–8 nucleotides in length. Each sample was PCR-amplified using different combinations of matching tags for each of three PCR replicates. PCR products were visualized with GelRed Nucleic Acid Stain (Biotum, Heyward, CA) on a 2% agarose gel against a 50 bp ladder. All negative controls appeared negative.

The PCR products were pooled into amplicon pools avoiding that PCR products carrying the same tag combination were pooled together. Pooling was based on gel band strength to approximate equimolar ratios. Negative controls were included at a similar volume as the PCR amplicons. Amplicon pools were converted into sequencing libraries following the protocol described in (Schnell et al., 2015) using a NEBNext DNA Library Prep Master Mix Set (#E6070; NEB, Ipswich, MA) and sequenced 230 bp paired-end on an Illumina MiSeq sequencing platform aiming for 20,000 reads per PCR replicate.

After sequencing, AdapterRemoval 2.1.7 (Lindgreen, 2012) was used to remove adaptors and low-quality reads. A customized Perl script was used to merge the paired reads. A modified version (<https://github.com/shyamsg/DAME>) of the tool kit DAME



(Zepeda-Mendoza, Bohmann, Carmona Baez, & Gilbert, 2016) was subsequently used to sort sequences by primer and tags and to filter sequences from each sample's PCR replicates so that only sequences present in minimum two of the three PCR replicates were retained, and singletons were removed (Alberdi, Aizpurua, Gilbert, & Bohmann, 2018). Operational taxonomic units (OTUs) were clustered at 97% similarity using Sumacust (Mercier, Boyer, Bonin, & Coissac, 2013). Each sample was normalized to a total of 50,000 sequences. In each sample, OTU sequences with a copy number of less than 0.1% of total (less than 500 copies) were discarded as these were assessed as likely false positives (OTU table with all OTUs and sequences before the 0.1% filtering is supplied as Table S4). Operational taxonomic units were identified through best matches in the Barcode of Life Data System (Ratnasingham & Hebert, 2007) or the NCBI database (<https://www.ncbi.nlm.nih.gov/>). OTUs were assigned to taxonomic order or family if it was >97.3% similar to a reference arthropod sequence. For consistency, OTU taxonomy and morphological identifications were not performed at lower taxonomic level than family. Operational taxonomic units not assigned to the phylum Arthropoda were not included in the study.

## 2.5 | Analysing results

To enable comparison of ct values between different qPCR runs, ct values were normalized against the mean ct value of the four positive controls (inter-run calibrators) (Bustin et al., 2010). Ct values, normalized ct values and number of OTUs were statistically analyzed using a paired t test. Venn diagrams and Jaccard similarity coefficients were calculated using Rstudio (Version 1.0.143) and the packages *vegan* and *VennDiagram*.

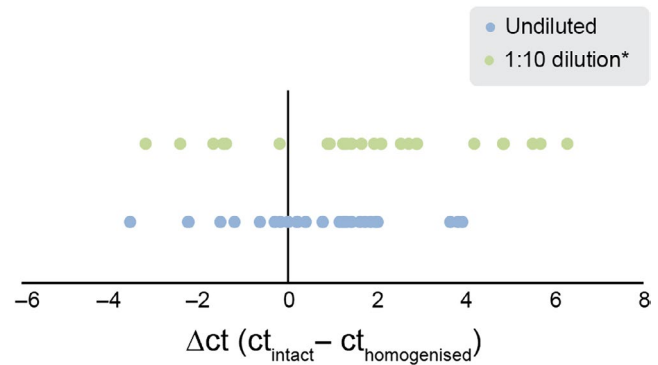
## 3 | RESULTS

### 3.1 | Non-destructive DNA extraction

Although initial physical homogenization of samples yielded higher DNA concentrations compared to samples that had not undergone homogenization prior to DNA extraction (Figure 2), they were also more strongly affected by PCR inhibitors (Figure S2). Notably, we also found that the OTU diversity recovered from intact samples was generally at least comparable to, if not better than, that from the homogenized samples with  $R^2$  values almost twice as high for intact samples compared to homogenized samples (Figure 3; Tables S2, S3). Furthermore, homogenized samples exhibited more variation (less consistency) in OTU numbers between the three purification approaches (phenol/chloroform inhibitor removal followed by automated purification, only manual purification, and only automatic purification) (Figure S3).

#### 3.1.1 | Digest volumes

We initially investigated whether the volume of purified digest affects the recovered OTU diversity. Within each of the three large



**FIGURE 2**  $\Delta ct$  values (calculated as the normalized  $ct_{intact} - ct_{homogenized}$ ) for undiluted (blue) and 1:10 diluted (green) sample extracts from the nine sample pairs with the three different purification approaches. Positive values indicate a higher DNA concentration for homogenized samples. \* indicate a significant difference (t test,  $p < 0.05$ ) in ct values between homogenized and intact samples

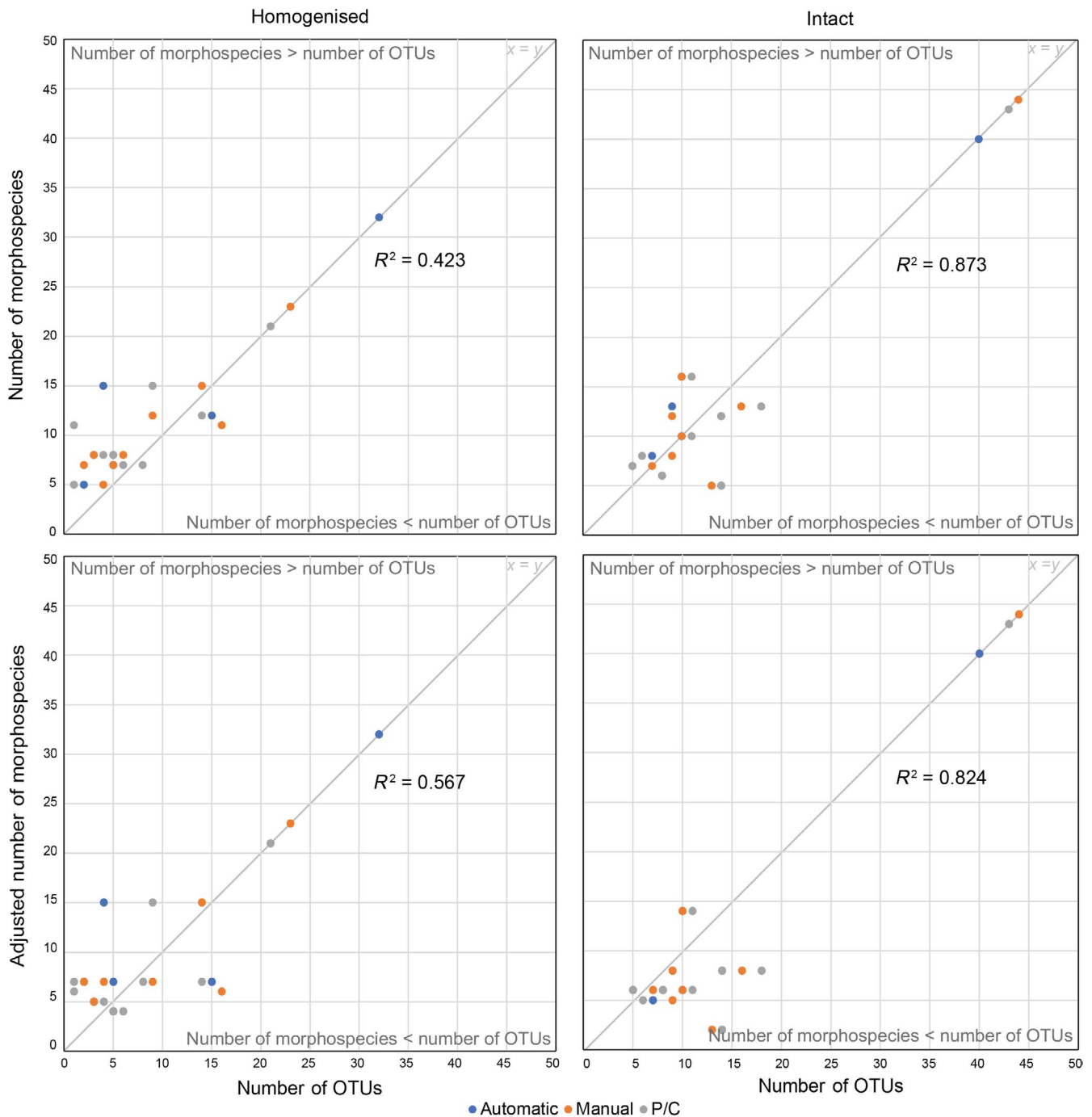
mock bulk samples, there was a consistent and large overlap of recovered OTUs irrespective of whether purification was carried out on 1,000 or 200  $\mu$ l digest volumes (Figure 4a). Importantly, purification of 1,000  $\mu$ l digest did not result in higher numbers of recovered OTUs within each of the three large mock bulk samples, meaning that purification of 200  $\mu$ l digest was enough to represent the overall OTU diversity (Figure 4b; Figure S1).

#### 3.1.2 | The effect of phenol/chloroform inhibitor removal

We subsequently explored the effect of purifying bulk arthropod sample digests with organic solvents (phenol/chloroform) to remove PCR inhibitors. While we found that this inhibitor removal yielded significantly ( $p < 0.01$ ) higher amounts of DNA (Figure 5), this did not affect the number of recovered OTUs ( $p > 0.05$ ). Indeed, samples that were not treated with phenol/chloroform performed equally well, if not better (Figure 6; Figures S4 and S5). We found no obvious difference in inhibitor levels between samples which were treated with phenol/chloroform and those that were not, and some samples still showed signs of PCR inhibitors even after the phenol/chloroform treatment (Figure S6).

#### 3.1.3 | Automated purification

Lastly, we explored the performance of automated purification of DNA extracts with a QIAcube robot versus manual purification. DNA yields were generally the same between the automated and manual purification (Figure S7), and no difference was observed with regard to PCR inhibitor levels (data not shown). Furthermore, the number of OTUs did not differ ( $p > 0.05$ ) between the two purification methods (Figure 7), and the Jaccard similarity coefficients were high showing similar OTU diversity outputs between the two methods (Figures S8 and S9).



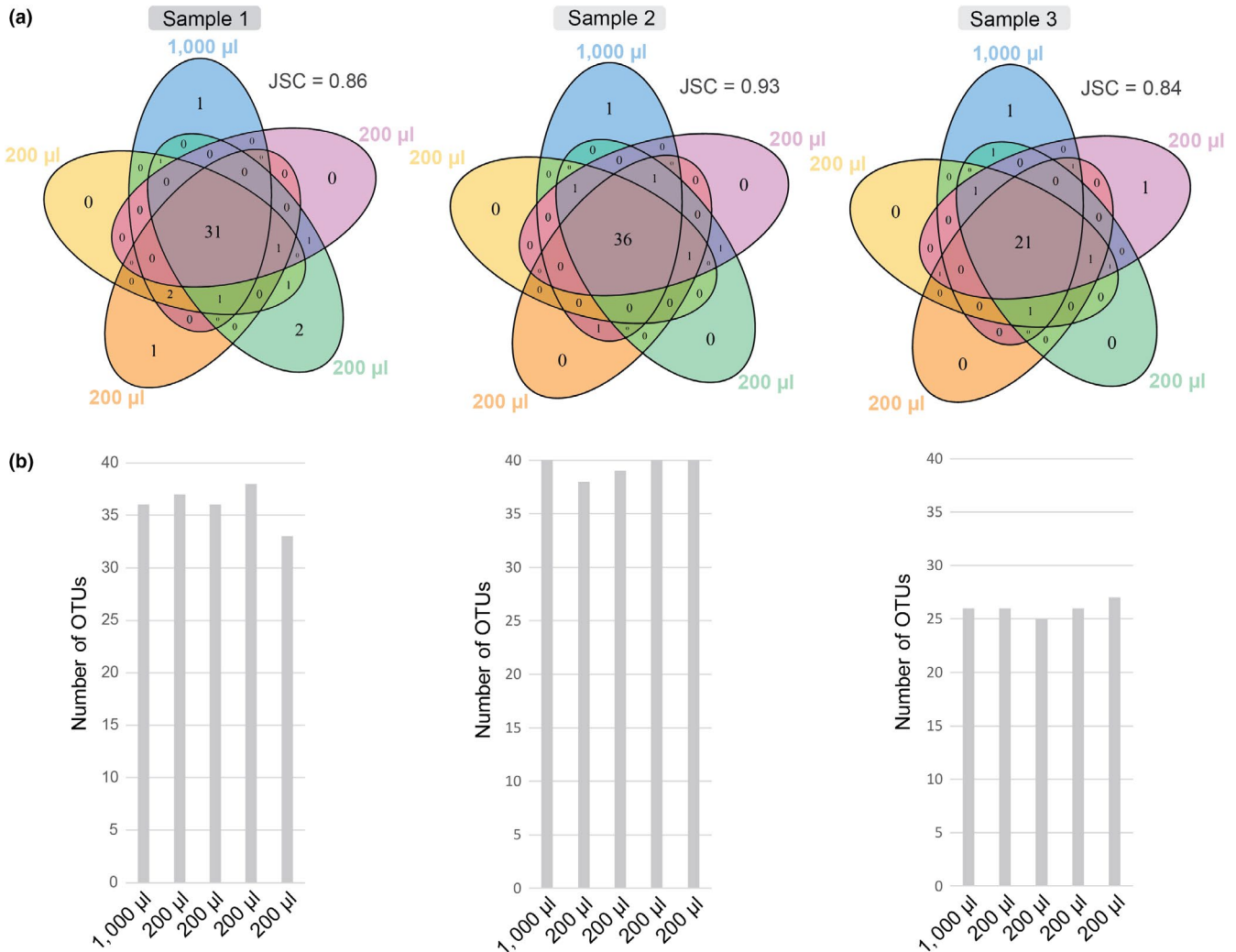
**FIGURE 3** Scatterplot for the number of operational taxonomic units (OTUs; x-axis) and observed number of morphospecies (y-axis) for the nine sample-pairs of homogenized (left) and intact (right) mock bulk arthropod samples. For these, three different approaches applied for the DNA purification are shown; automatic purification using a QIAcube robot (Automatic), manual purification (Manual), and phenol/chloroform inhibitor removal followed by automatic purification (P/C). The two top plots include all observed number of morphospecies whereas the bottom plots include an adjusted number of morphospecies in relation to the known biases of the used primer set (Tables S2 and S3). Data points above the  $x = y$  line indicates a higher number of morphospecies than the number of retained OTUs.  $R^2$  values for an intercept at (0,0) are shown

## 4 | DISCUSSION

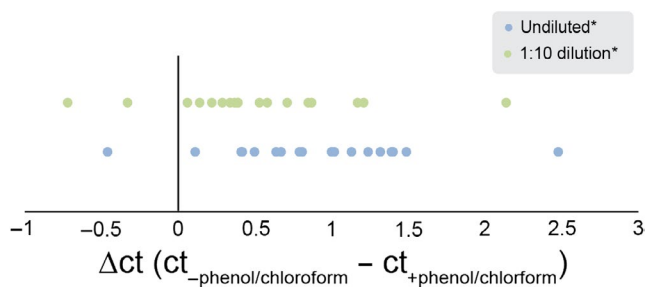
The principal aim of this study was to explore whether it is possible to increase the efficiency of DNA extractions from keeping bulk caught arthropod samples unsorted and without losing taxonomic

information. Our results demonstrate how steps relating both to sample pre-processing and DNA extraction can be simplified without incurring significant loss of biodiversity information.

First, we demonstrate that it is not necessary to physically homogenize the arthropods into a “soup”, and we believe that this offers



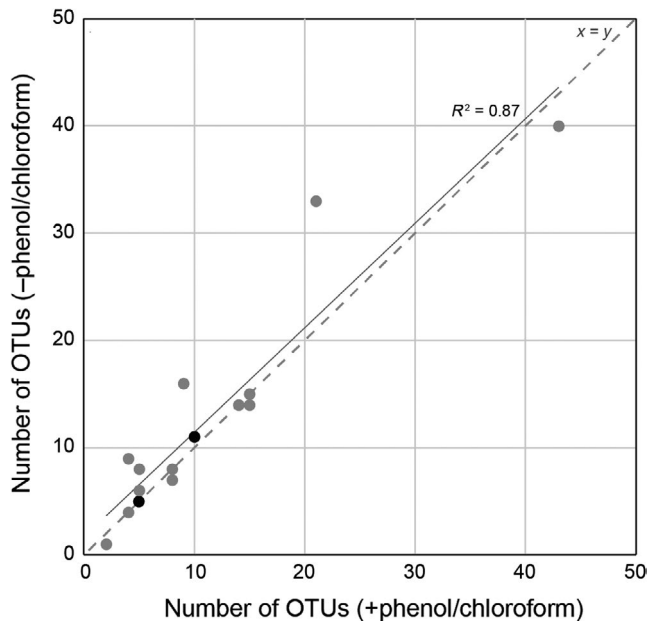
**FIGURE 4** Assessing the effect of purifying different digest volumes from arthropod bulk samples. For each of three large mock bulk arthropod samples, 1000 µl digest and four replicates of 200 µl digest were purified, resulting in five DNA extracts from each sample on which metabarcoding was carried out. (a) Overlap and (b) number of operational taxonomic units (OTUs) resulting from purifying 1000 µl and each of four replicates of 200 µl digest. The mean Jaccard similarity coefficient (JSC) indicates the similarity between the five DNA extracts. The higher the JSC, the higher similarity



**FIGURE 5**  $\Delta ct$  values (calculated as the normalized  $ct_{-phenol/chloroform} - ct_{+phenol/chloroform}$ ) for undiluted (blue) and 1:10 diluted (green) sample extracts from the nine sample pairs with the three different purification approaches. Positive values indicate a higher DNA concentration for samples treated with phenol/chloroform. \*Indicates a significant difference ( $t$  test,  $p < 0.01$ ) in  $ct$  values between samples treated with phenol/chloroform and those that were not

considerable potential gains. This is not only because eliminating the physical homogenization step retains the information represented by the specimens' morphology, but also because sample pre-processing rapidly becomes impractical and costly in man-hours when handling large numbers of samples. Furthermore, the incorporation of additional processing steps increases the risk of cross-contamination between samples. Perhaps the only disadvantage of not homogenising the samples, is the lower levels of DNA (Figure 2). Whether this is actually a disadvantage, however, is unclear and we hypothesize that DNA amplified from non-homogenized samples better reflects the overall community composition because larger individuals will not contribute with an equally larger DNA release when kept intact as opposed to being homogenized. This is because homogenized specimens will contribute in proportion to their full biomass (an approximately cubic relationship to their cross-sectional radius), whereas when left unhomogenized the DNA released is a function of their surface area, thus



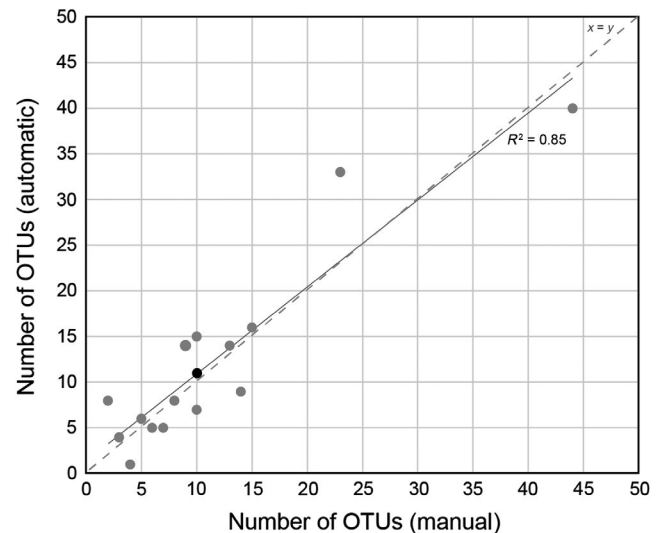


**FIGURE 6** Comparison of the number of operational taxonomic units (OTUs) recovered from samples for which the DNA extraction protocol included a phenol/chloroform inhibitor removal step (x-axis) and where it did not (y-axis). Specifically, DNA digests from 18 bulk arthropod samples were treated with a phenol/chloroform inhibitor removal step and compared against the same DNA digest, which did not undergo phenol/chloroform inhibitor removal. Black marks indicate two data points in the same place. The linear regression line is depicted as a solid line with associated  $R^2$  value

proportional to the square of their cross-sectional radius. Whether this is the case requires further study, although it might explain the inconsistent OTU diversity recovered within the same homogenized samples but purified in three different ways (Figures S3–S5, S8, S9). It should be noted that a simple sorting of specimens within bulk arthropod samples according to their size (as in e.g. Elbrecht et al., 2017) followed by non-destructive extraction of each size group, might further reduce bias caused by different specimen sizes, but will also increase the number of samples substantially.

Another non-destructive approach of assessing biodiversity was reported by Hajibabaei et al. (2012), who extracted invertebrate DNA from their preservative ethanol thus keeping individual sample constituents intact. However, this approach might not be suitable for detecting the overall diversity in bulk arthropod samples as it failed to detect a large proportion of the arthropod species in a known bulk arthropod sample (Erdozain et al., 2019; Linard, Arribas, Andújar, Crampton-Platt, & Vogler, 2016).

Second, we demonstrate that while extraction of DNA from bulk-sampled arthropods typically requires large volumes (e.g. multiple ml) of digestion buffer unless they are first subsampled, purification of only a sub-fraction of the digestion buffer maintains the overall OTU diversity (Figure 4). This in turn renders savings in terms of both time and economic costs, and further makes automatic handling more feasible as a fixed digest volume is often the only option.



**FIGURE 7** Number of operational taxonomic units (OTUs) resulting from metabarcoding analyses of DNA extract digest from 18 bulk arthropod samples, which were all both purified on a QIAcube robot (y-axis) and manually purified (x-axis). Black data points indicate two data points in the same place. One homogenized sample purified on a QIAcube robot failed to PCR amplify and was not used. The linear regression line is depicted as a solid line and the  $R^2$  value included

Third, we demonstrate that although organic solvents such as phenol/chloroform have been used in some previous studies on arthropod genetics (Gilbert et al., 2007; Zhou et al., 2013), possibly with the aim of ensuring potential inhibitory substances are purified out of DNA extracts (Hunt, 1997; Sperling, Anderson, & Hickey, 1994), their use offered no significant benefit often resulting in the same or lower numbers of OTUs compared with their omission (Figure 6; Table S2–S3). This likely indicates that the occurrence of PCR inhibitors in DNA extracts from bulk arthropod samples are equally well accounted for through dilution of DNA extracts and given both the toxicity of the organic solvents, and the manual labor required to use them, their omission is a welcome step.

Lastly, we show that automated bulk arthropod digest purification on the QIAcube platform is an effective way to optimize their processing, as it returns similar results to manual purification (Figure 7), while concomitantly offering savings in manual labor, and almost certainly improving consistency and reducing the risk of human error and contamination. We acknowledge that the QIAcube robot used in this study only has a 12-sample throughput and therefore cannot be seen as high-throughput processing. The benefits obtained from automated sample handling can, however, also be obtained from other platforms such as QIA Symphony (Qiagen) or Opentrons (Opentrons, New York, USA) with higher sample-outputs.

Naturally, despite these successes we would be amiss if we were not to highlight the fact that our results are based upon a relatively limited dataset, and thus come with several caveats. First, we highlight that our results are based on metabarcoding. While currently the most popular approach in molecular biodiversity studies, the potential of alternate approaches are now being explored, such as

shotgun and target capture enrichment sequencing as a means of characterizing the taxonomic diversity of complex arthropod communities (e.g. Tang et al., 2015; Zhou et al., 2013). While utilization of such approaches is unlikely to change the results, studies that directly validate this will be welcomed. Metabarcoding also relies on PCR amplification which might introduce primer bias (Elbrecht & Leese, 2015; Piñol, Mir, Gomez-Polo, & Agustí, 2015). The primer set used in this study, ZBJ-ArtF1c and ZBJ-ArtR2c (Zeale et al., 2011) is known to have lesser affinity toward, for example Hymenoptera and Coleoptera (Brandon-Mong et al., 2015; Clarke, Soubrier, Weyrich, & Cooper, 2014). In this study, all samples are, however, influenced by the same primer bias, and we found no difference in results when taking the potential biases into account (Figure 3; Tables S2, S3).

Second, we also acknowledge that the complexity of the arthropod samples analysed here is moderate, as our mock samples contain no more than 48 different taxa (Tables S2, S3). Other studies will perhaps face far richer bulk samples – for example, one single Malaise trap in New Zealand was reported to catch an average of 3,800 individual arthropods per month across a year of sampling (Moeed & Meads, 1987), while a Malaise trap deployed in Costa Rica collected more than a thousand arthropod specimens during a 7-day period, resulting in more than 350 OTUs (Gibson et al., 2014). While an increase in community complexity is unlikely to affect either organic solvent or automated purification results, and as indeed the current extraction protocol may even perform better on non-homogenized samples for the reasons outlined above, there is certainly a need for future studies to explore whether purification of subsamples of digests is equally effective when applied to such complex mixtures.

Lastly, we highlight that while our method is certainly less destructive than homogenization, given that DNA is being extracted, it is clearly not a completely non-destructive method. The DNA principally derives from the internal soft tissue (Gilbert et al., 2007), with the external features left largely intact, but even so the degree of “non-destructiveness” to these features will be taxon-dependent, and more robust specimens (e.g. those with a more sclerotized exoskeleton such as most beetles) will be better suited than others for subsequent analysis or inclusion in museum collections post extraction. Having said that, in separate pilot experiments we have succeeded in extracting DNA using the protocol introduced here on a range of more fragile specimens including flies, and subsequently included the specimens into museum collections afterwards (unpublished). Therefore, while promising, we recommend that potential users may wish to undertake pilot tests before fully adopting the method on valuable specimens. One should also note that when scaling up, a large volume of digest buffer will be needed. We have estimated a cost of around 3.8\$ per sample (using 10 ml buffer) and increasing further for larger bulk samples. The benefits of the method presented here, such as not having to sort the samples (therefore fewer samples for the downstream work), reduced handling time and the non-destructive approach outweighs the potential small increase in extraction expenses.

We have demonstrated how steps related to both sample pre-processing and DNA extraction of bulk arthropod samples can be simplified and automated without incurring significant loss of biodiversity

information. This workflow is efficient, cost-effective, reduces risk of cross-contamination, and importantly, leaves specimens intact. We believe this has the potential to change the way we utilize and value DNA-based biodiversity studies of arthropods in the future.

## ACKNOWLEDGMENTS

We are grateful for the permits provided by Tanzania National Parks (TANAPA), the Tanzania Commission for Science and Technology (COSTECH: No. 2013-242-ER-2012-147 to TP and 2014-138-ER-2012-147 to MN) and the Tanzania Wildlife Research Institute (TAWIRI). We also thank Richard M. L. Laizzer and Aloyce Mwakisoma for their invaluable assistance in the field, and the Udzungwa Ecological Monitoring Centre (UEMC) and its staff for logistical support. Furthermore, we thank the Danish National High-Throughput DNA Sequencing Centre for sequencing. We are also grateful to Nikolaj Scharff for his help with samples, logistics and read through. KB was funded by the Independent Research Fund Denmark (DFF 5051-00140), TP's field work was supported by the Carlsberg Foundation (grant number 2012\_01\_0504) and MN received funding from Dr. Bøje Benzons support grant and from the EvoGenomics MSc grant.

## CONFLICT OF INTEREST

The authors declare that they have no competing interests.

## AUTHORS' CONTRIBUTIONS

MN, MTPG, and KB designed the study. MN and TP conducted morphospecies assessment. MN carried out laboratory work and data analysis. MN and KB wrote the first draft and all authors contributed to revisions and accepted the final version.

## DATA ACCESSIBILITY

Sequencing data and information on tag and primer combinations in libraries are available at <http://www.doi.org/10.17894/ucph.356ebf59-2d6b-4550-a897-fb940778a4f9>

## ORCID

Martin Nielsen  <https://orcid.org/0000-0002-3718-526X>

M. Thomas P. Gilbert  <https://orcid.org/0000-0002-5805-7195>

Thomas Pape  <https://orcid.org/0000-0001-6609-0609>

Kristine Bohmann  <https://orcid.org/0000-0001-7907-064X>

## REFERENCES

- Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution/British Ecological Society*, 9, 134–147. <https://doi.org/10.1111/2041-210X.12849>

- Baird, D. J., & Hajibabaei, M. (2012). Biomonitoring 2.0: A new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, 21(8), 2039–2044. <https://doi.org/10.1111/j.1365-294x.2012.05519.x>
- Basset, Y., Cizek, L., Cuénoud, P., Didham, R. K., Guilhaumon, F., Missa, O., ... Leponce, M. (2012). Arthropod diversity in a tropical forest. *Science*, 338(6113), 1481–1484. <https://doi.org/10.1126/science.1226727>
- Beng, K. C., Tomlinson, K. W., Shen, X. H., Surget-Groba, Y., Hughes, A. C., Corlett, R. T., & Slik, J. W. F. (2016). The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land-use change in the tropics. *Scientific Reports*, 6, 24965. <https://doi.org/10.1038/srep24965>
- Binladen, J., Gilbert, M. T. P., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R., & Willerslev, E. (2007). The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE*, 2(2), e197. <https://doi.org/10.1371/journal.pone.0000197>
- Brandon-Mong, G.-J., Gan, H.-M., Sing, K.-W., Lee, P.-S., Lim, P.-E., & Wilson, J.-J. (2015). DNA metabarcoding of insects and allies: An evaluation of primers and pipelines. *Bulletin of Entomological Research*, 105(6), 717–727. <https://doi.org/10.1017/S0007485315000681>
- Bustin, S. A., Beaulieu, J.-F., Huggett, J., Jaggi, R., Kibenge, F. S. B., Olsvik, P. A., ... Toegel, S. (2010). MIQE precis: Practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. *BMC Molecular Biology*, 11, 74. <https://doi.org/10.1186/1471-2199-11-74>
- Clarke, L. J., Soubrier, J., Weyrich, L. S., & Cooper, A. (2014). Environmental metabarcodes for insects: In silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, 14(6), 1160–1170. <https://doi.org/10.1111/1755-0998.12265>
- Crampton-Platt, A., Timmermans, M. J. T. N., Gimmel, M. L., Kutty, S. N., Cockerill, T. D., Vun Khen, C., & Vogler, A. P. (2015). Soup to Tree: The phylogeny of beetles inferred by mitochondrial metagenomics of a Bornean rainforest sample. *Molecular Biology and Evolution*, 32(9), 2302–2316. <https://doi.org/10.1093/molbev/msv111>
- Elbrecht, V., & Leese, F. (2015). Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass–Sequence Relationships with an Innovative Metabarcoding Protocol. *PLoS ONE*, 10(7), e0130324. <https://doi.org/10.1371/journal.pone.0130324>
- Elbrecht, V., Peinert, B., & Leese, F. (2017). Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. *Ecology and Evolution*, 7(17), 6918–6926. <https://doi.org/10.1002/ece3.3192>
- Erdozain, M., Thompson, D. G., Porter, T. M., Kidd, K. A., Kreutzweiser, D. P., Sibley, P. K., ... Hajibabaei, M. (2019). Metabarcoding of storage ethanol vs. conventional morphometric identification in relation to the use of stream macroinvertebrates as ecological indicators in forest management. *Ecological Indicators*, 101, 173–184. <https://doi.org/10.1016/j.ecolind.2019.01.014>
- Gibson, J., Shokralla, S., Porter, T. M., King, I., van Konynenburg, S., Janzen, D. H., ... Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasyntematics. *Proceedings of the National Academy of Sciences of the United States of America*, 111(22), 8007–8012. <https://doi.org/10.1073/pnas.1406468111>
- Gilbert, M. T. P., Moore, W., Melchior, L., & Worobey, M. (2007). DNA extraction from dry museum beetles without conferring external morphological damage. *PLoS ONE*, 2(3), e272. <https://doi.org/10.1371/journal.pone.0000272>
- Gillett, C. P. D. T., Crampton-Platt, A., Timmermans, M. J. T. N., Jordal, B. H., Emerson, B. C., & Vogler, A. P. (2014). Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). *Molecular Biology and Evolution*, 31(8), 2223–2237. <https://doi.org/10.1093/molbev/msu154>
- Hajibabaei, M., Spall, J. L., Shokralla, S., & van Konynenburg, S. (2012). Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecology*, 12, 28. <https://doi.org/10.1186/1472-6785-12-28>
- Hunt, G. J. (1997). Insect DNA Extraction Protocol. In M. R. Michelli & R. Bova (Eds.), *Fingerprinting methods based on arbitrarily primed PCR* (pp. 21–24). Berlin, Heidelberg: Springer.
- Kocher, A., Gantier, J.-C., Gaborit, P., Zinger, L., Holota, H., Valiere, S., ... Murienne, J. (2017). Vector soup: High-throughput identification of Neotropical phlebotomine sand flies using metabarcoding. *Molecular Ecology Resources*, 17(2), 172–182. <https://doi.org/10.1111/1755-0998.12556>
- Linard, B., Arribas, P., Andújar, C., Crampton-Platt, A., & Vogler, A. P. (2016). Lessons from genome skimming of arthropod-preserving ethanol. *Molecular Ecology Resources*, 16(6), 1365–1377. <https://doi.org/10.1111/1755-0998.12539>
- Lindgreen, S. (2012). AdapterRemoval: Easy cleaning of next-generation sequencing reads. *BMC Research Notes*, 5, 337. <https://doi.org/10.1186/1756-0500-5-337>
- Liu, S., Li, Y., Lu, J., Su, X., Tang, M., Zhang, R., ... Zhou, X. (2013). SOAPBarcode: Revealing arthropod biodiversity through assembly of Illumina shotgun sequences of PCR amplicons. *Methods in Ecology and Evolution/British Ecological Society*, 4(12), 1142–1150. <https://doi.org/10.1111/2041-210x.12120>
- Mercier, C., Boyer, F., Bonin, A., & Coissac, E. (2013). SUMATRA and SUMACLUSt: fast and exact comparison and clustering of sequences. In Programs and Abstracts of the SeqBio 2013 workshop. Abstract (pp. 27–29).
- Moeed, A., & Meads, M. J. (1987). Seasonality of arthropods caught in a Malaise trap in mixed lowland forest of the Orongorongo Valley, New Zealand. *New Zealand Journal of Zoology*, 14(2), 197–208. <https://doi.org/10.1080/03014223.1987.10422990>
- Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How many species are there on Earth and in the ocean? *PLoS Biology*, 9(8), e1001127. <https://doi.org/10.1371/journal.pbio.1001127>
- Morinière, J., Cancian de Araujo, B., Lam, A. W., Hausmann, A., Balke, M., Schmidt, S., ... Haszprunar, G. (2016). Species identification in malaise trap samples by DNA barcoding based on NGS technologies and a scoring matrix. *PLoS ONE*, 11(5), e0155497. <https://doi.org/10.1371/journal.pone.0155497>
- Murray, D. C., Coghlan, M. L., & Bunce, M. (2015). From benchtop to desktop: Important considerations when designing amplicon sequencing workflows. *PLoS ONE*, 10(4), e0124671. <https://doi.org/10.1371/journal.pone.0124671>
- Noriega, J. A., Hortal, J., Azcarate, F. M., Berg, M. P., Bonada, N., Briones, M. J. I., ... Santos, A. M. C. (2018). Research trends in ecosystem services provided by insects. *Basic and Applied Ecology*, 26, 8–23. <https://doi.org/10.1016/j.baae.2017.09.006>
- Piñol, J., Mir, G., Gomez-Polo, P., & Agustí, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources*, 15(4), 819–830. <https://doi.org/10.1111/1755-0998.12355>
- Rainio, J., & Niemelä, J. (2003). Ground beetles (Coleoptera: Carabidae) as bioindicators. *Biodiversity and Conservation*, 12(3), 487–506.
- Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The barcode of life data system ([www.barcodinglife.org](http://www.barcodinglife.org)). *Molecular Ecology Notes*, 7, 355–364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x>
- Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated—Reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, 15(6), 1289–1303. <https://doi.org/10.1111/1755-0998.12402>
- Schrader, C., Schielke, A., Ellerbroek, L., & John, R. (2012). PCR inhibitors—Occurrence, properties and removal. *Journal of Applied Microbiology*, 113(5), 1014–1026. <https://doi.org/10.1111/j.1365-2672.2012.05384.x>

- Sperling, F. A. H., Anderson, G. S., & Hickey, D. A. (1994). A DNA-based approach to the identification of insect species used for postmortem interval estimation. *Journal of Forensic Sciences*, 39(2), 1361–1371. <https://doi.org/10.1520/JFS13613J>
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 21(8), 2045–2050. <https://doi.org/10.1111/j.1365-294X.2012.05470.x>
- Tang, M., Hardman, C. J., Ji, Y., Meng, G., Liu, S., Tan, M., ... Yu, D. W. (2015). High-throughput monitoring of wild bee diversity and abundance via mitogenomics. *Methods in Ecology and Evolution / British Ecological Society*, 6(9), 1034–1043. <https://doi.org/10.1111/2041-210X.12416>
- Thomas, J. A., Telfer, M. G., Roy, D. B., Preston, C. D., Greenwood, J. J. D., Asher, J., ... Lawton, J. H. (2004). Comparative losses of British butterflies, birds, and plants and the global extinction crisis. *Science*, 303(5665), 1879–1881. <https://doi.org/10.1126/science.1095046>
- Wang, W. Y., Srivathsan, A., Foo, M., Yamane, S. K., & Meier, R. (2018). Sorting specimen-rich invertebrate samples with cost-effective NGS barcodes: Validating a reverse workflow for specimen processing. *Molecular Ecology Resources*, 18(3), 490–501. <https://doi.org/10.1111/1755-0998.12751>
- Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution / British Ecological Society*, 3, 613–623. <https://doi.org/10.1111/j.2041-210X.2012.00198.x>
- Zeale, M. R. K., Butlin, R. K., Barker, G. L. A., Lees, D. C., & Jones, G. (2011). Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Molecular Ecology Resources*, 11(2), 236–244. <https://doi.org/10.1111/j.1755-0998.2010.02920.x>
- Zepeda-Mendoza, M. L., Bohmann, K., Carmona Baez, A., & Gilbert, M. T. P. (2016). DAME: A toolkit for the initial processing of datasets with PCR replicates of double-tagged amplicons for DNA metabarcoding analyses. *BMC Research Notes*, 9, 255. <https://doi.org/10.1186/s13104-016-2064-9>
- Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X. U., Zhou, L., ... Huang, Q. (2013). Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, 2(1), 4. <https://doi.org/10.1186/2047-217X-2-4>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Nielsen M, Gilbert MTP, Pape T, Bohmann K. A simplified DNA extraction protocol for unsorted bulk arthropod samples that maintains exoskeletal integrity. *Environmental DNA*. 2019;1:144–154. <https://doi.org/10.1002/edn3.16>